Data Analysis: What factors contribute to being good at soccer?

Eniola Osineye

Walsh University

Abstract

Soccer is the world’s sport - it is the most played and viewed sport around the world. With a soccer game every day of the week, the sport amasses a large amount of interest. The Fédération Internationale de Football Association (FIFA) is the governing body of international associations and organizes tournaments involving many countries. The most viewed and the loved tournament is the FIFA World Cup, where every four years, the best countries around the world meet in a host country and battle to be crowned the best in the world. With many players and leagues around the world, it is hard to determine who is the best player and what makes a player good enough to be amongst the best. This statistical analysis will look at the best players from the top five leagues in Europe and use data such as their weight, height, goals, and assists to determine what factors contribute to a player winning the most trophies like the pinnacle trophy, the FIFA World Cup. The main variable that will be focused on is the weight of soccer player. Statistical analysis such as frequency distributions contingency tables, hypothesis, and ANOVA test will be used to answer the question of what factors contribute to being good at soccer?

Data Analysis: What factors contribute to being good at soccer?

In today's world, the amount of data available to us is immense and trying to arrange it can be overwhelming. With the help of statistical analysis, one can use this data to answer important questions that can be used to win arguments, for broadcast reporting or betting. As one of many soccer fans around the world, one becomes interested in the number of goals, assists, minutes, and trophies won by players and how this contributes to their status as legends. For the analysis, information on the best soccer players based on their status in the footballing world was composed using several soccer databases that are listed in the references. Every player in the population, in their regard, is currently considered a soccer legend and one of the best players at their respective positions. So, the population is composed of players from every outfield position, attackers, midfielders, and defenders. Along with the soccer player’s positions, sixteen other variables were retrieved. These variables included, weight, height, nationality, matches played, minutes played, goals, assists, league trophies, individual awards, champions leagues trophies, and seasons played. After completing descriptive statistics for each variable of the population; the mean, median, mode, standard deviation, and variance were given. With the mean height given at 179.53 cm, it can be concluded that the best soccer players are of average height and with the mean weights of 75.63 Kg. Therefore, it can be concluded that the best soccer player is below the average weight of men with the age group of 20 – 30. This is expected as athletes are expected to be fit and toned.

Probabilities are used to determine how likely something is going to happen. Regarding the population, the likelihood of the best soccer player being an attacker is 46.7%, whiles the likelihood of being a midfielder is 26.7% and a defender is 26.7%. The data shows that 43% of the population has scored less than 38 goals, 53% of the population is between the weight of 68Kg and 75Kg, 50% of the population has played many matches between 323 to 402, and 65.26% of the population have played 23,185 – 35,183 minutes across their career. Moreover, 33.33% of players in the population have provided 5 – 44 assists while 43.33%, 20%, and 3.33% of players have provided 45- 84, 85 – 124, and 165 – 204 assists respectively. It can be concluded that many players in the population have provided less than 84 assists which is significantly less than the highest number of assists provided of 178. Therefore, one can conclude that the best players in the world provide between the range of 5 – 84 assists throughout their careers. Moreover, regarding preferred foot, right-footed players scored 61.27% of the population’s goals and provided 58.44% of the population’s assists. This suggests that many of the best players are right-footed, and this is supported by 70% of the population being right-footed. However, the analysis will disregard the data about a preferred foot because it is a known fact that there are a significant number of more right-handed and right-footed people compared to left-handed and left-footed people in the world, therefore, being right-footed does not necessarily mean that you will be a good soccer player as a lot of players are right-footed in general.

Marginal probability tells us the likelihood of an event happening regardless of the outcome of another variable. So, in this analysis, it tries to answer the question of whether winning the FIFA World Cup influences whether a player is considered one of the best in the world. Based on the results, the marginal probability of a player in the population that has won the FIFA World Cup is 33%. This leads to the conclusion that most players have not won the FIFA World Cup and are still considered one of the best players in the world. Therefore, a FIFA World cup trophy is not a contributing factor to being one of the best players in the World.

Moreover, the analysis used normal distributions to try and answer whether the weight of a player attributes to being considered one of the best players in the world. The analysis assumed that players that weigh less are more fit and faster than their opponents. The results showed that the probability that a player weighs less than 70kg in a normal distribution is 25%. Also, the probability that a player weighs between 70kg and 80kg in a normal distribution is 46%. From the results, the initial assumption is wrong. The results point to the theory that being lighter does not necessarily mean that the player will become one of the best in the world.

With the confidence interval test that was done, the analysis took a sample of 9 players and wanted to know with a confidence interval of 95% what the mean weight of the sample was. The results concluded that the mean weight is 68Kg with a large margin of error of 7.01Kg. This test did not provide much of an insight as it showed me that the sample’s mean weight had a large margin of error.

Using hypothesis testing, the analysis proposed the alternative hypothesis that the soccer players at the mean weight of the population were more likely to be amongst the best soccer players. A sample of 14 players was taken and the results showed that the null hypothesis must be rejected, and it would be wrong to assume that the mean amount of soccer players' weight is 75.63kg. Moreover, another hypothesis test was done to determine whether a player greater than the mean weight is more likely to be the best player. The results concluded that it should not be assumed that the mean weight is greater than or equal to the mean weight.

Finally, an ANOVA test was performed to determine from three samples of five from the population, whether the population means are equal. The results concluded that the null hypothesis that the three-population means are equal must be rejected. This means across the three samples, each sample’s mean is significantly different. Another ANOVA test was performed to determine if there is a difference in the mean of three samples of 10 players of the total trophies won by each player including league trophies, cups, and individual trophies. The results concluded that one should accept the null hypothesis, that the mean of three samples of trophies won by players in the population are the same. Therefore, on average to be one of the best soccer players, the player should win 22 trophies. However, whiles calculating this ANOVA test, there were a few outliers and with those outliers removed, the same conclusion is reached.

In summation, with all these analyses, each variable was examined, and it was concluded that the variable either contributes to being one of the best soccer players or not. The main factor that was focused on was weight and it can be concluded the average weight of the best soccer players is 75.63Kg. Moreover, many of the best players in the world have played between 323 to 402 games, and 23,185 to 35, 183 minutes. Besides, most of the best soccer players are of the height of 179.53cm and the majority have scored less than 38 goals and assisted less than 84 assists. Moreover, the preferred foot was disregarded as a factor due to the large ratio of right-footed players in the world and the data supporting this fact. Also, to be one of the best soccer players in the word, the player does not necessarily need to win a FIFA World Cup, the most coveted price in soccer, but should win an average of 22 trophies. Becoming a soccer player is extremely difficult as the number of total soccer players across the world dwarfs the amount of those players who play professionally. Therefore, this analysis tries to answer what factors contribute to being one of the best soccer players in the world. In hindsight, the analysis should have increased the amounts of data in the population to reduce the margin of error for many of the analyses.

References

Datopian. (n.d.). Football Data. Retrieved from https://datahub.io/collections/football

Football Transfers, Rumours, Market Values, News and Statistics. (n.d.). Retrieved from https://www.transfermarkt.us/

GiveMeSport. (2019, December 28). Ranking the 100 greatest footballers of the decade (2010-2020). Retrieved from https://www.givemesport.com/1533225-ranking-the-100-greatest-footballers-of-the-decade-20102020

Mathien, H. (2016, October 23). European Soccer Database. Retrieved from https://www.kaggle.com/hugomathien/soccer

There are 391 580 players in the database of footballdatabase.eu, covering the whole world of football. (n.d.). Retrieved from https://www.footballdatabase.eu/en/players